Upskilling Together: How Peer-interaction Influences Speaking-skills Development Online

Raiyan Abdul Baten
Electrical and Computer Engineering
University of Rochester
Rochester, United States
rbaten@ur.rochester.edu

Famous Clark
Computer Science
University of Rochester
Rochester, United States
fclark2@u.rochester.edu

Mohammed (Ehsan) Hoque
Computer Science
University of Rochester
Rochester, United States
mehoque@cs.rochester.edu

***Abstract*—We explore the characteristics and values of online peer-interactions in developing a fundamental soft-skill such as speaking. 60 participants recorded speech videos on 5 job-interview prompts and exchanged comments and performance ratings with their peers. We find that both (i) receiving suggestions for improvement ('tips') and (ii) having access to peers with better average ratings than one's own correspond to performance improvement (p<0.001 for both). Using linguistic features (e.g., emotions, personality, sentiment), we are able to classify tips from non-tip comments (AUC 0.89). Linguistic features from the received comments and the average ratings of one's peers incrementally improve the prediction of one's future ratings, showing the simultaneous importance of the two peer-learning sources. Qualitative analysis reveals dyadic and community-level peer-influence factors: context-driven feedback, first-hand demonstration, empathetic support, acknowledgment, opinion diversity, sense of community and comfort in interaction. These insights inform the building of intelligent human-machine symbiosis systems for speaking-skills development.**

***Index Terms*—Speaking-skills, Online peer learning**

I. Introduction

Across the global economy, employers have consistently highlighted competence in oral communication as a key soft-skill needed in today's workplace [1]–[3]. To meet the skill's ever-growing demand, its training efforts have spanned expert coaching to AI-driven feedback systems over the years. Entities such as Toastmasters [4], Own the Room [5] etc. offer expert guidance on speaking-skills, yet can face issues with scalability, on-demand availability and affordability. Recent developments in automated tools have enabled people to receive objective feedback on their speeches ubiquitously and cheaply [6]–[9]. But a lack of contextual understanding and personalization mark a bottleneck in this training form [10].

As we continue to make rapid progress in AI, there is a growing consensus that the purpose of AI should be to enhance human ability, rather than replace it [11]. We emphasize the idea of human-machine symbiosis as a middle ground in the speaking-skills training spectrum—by incorporating subjective feedback from online peer-learners to the skill development loop [7], [8]. While peer-learners are not trained experts, recent advances have nonetheless shown promise. For instance, it has been reported that participants gradually improve with and perform closer to the peers they interact with in speaking-skill learning communities [12]. If AI is to support such subjective exchanges intelligently, it becomes essential to first understand the various ways peer-interaction can aid the speaking-skill development process. In this paper, we shed light on this less explored human-centric upskilling opportunity. This leads to our first research question:

R1. *How does peer-interaction influence speaking performance?*

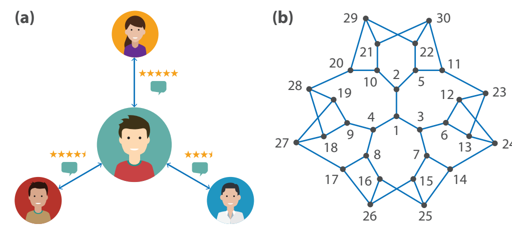

Fig. 1. (a) The participants exchanged comments and performance ratings with their peers. We analyze how the interactions correspond to speaking-skills development. (b) Network structure used in the study. For example, ID 1 interacts only with IDs $2, 3$ and 4. All edges are bidirectional.

We collected a dataset where the participants ($N = 60$) recorded video responses to 5 job interview prompts in 10 days. Each of them exchanged comments and ratings with 3 fixed peers, on specified speaking-skill attributes (Figure 1a). We analyze the effects of two major sources of peer-influence: receiving comments, and observing peers' performances. We find that videos that received suggestions for improvement (referred to as 'tips') had a significant increase in ratings in their following prompts ($p < 0.001$). We further find that when a participant has access to peers with an average rating better than his/her own rating, it corresponds with an improvement in his/her performance ($p < 0.001$). Peer-learning literature typically focuses on written or face-to-face exchanges in educational contexts, and not so much on such video-based learning among peers that we show evidence for [13]–[16].

To understand the characteristics of the improvement suggestions or 'tips' better, we ask our second research question:

R2. *Can we classify which comments have suggestions for improvement and which do not using linguistic cues?*

This work was supported by National Science Foundation Award IIS-1750380, and a Google Faculty Research Award.

For each comment, we extract 17 scores on basic emotions, social personality, language characteristics and sentiment (referred to as 'Affective features'). We demonstrate that the tips are typically more tentative, less positive and less extraverted in their tone than the non-tips. We also show that the affective features can classify tips from non-tips (AUC 0.89). This opens up the possibility of automatically sensing suggestion comments in speaking-skill development platforms. Such sensing can enable the development of intelligent interaction and intervention strategies to better support peer-learning.

Using a linear model, we find that the affective features from received comments and the peers' average ratings incrementally improve the prediction of one's future ratings. This shows the concurrent effects of both of the peer-learning sources we study. Through complementary qualitative analysis, we attempt to identify the human values AI cannot produce yet:
R3. *What are the dyadic and community-level peer-influence factors that add value to speaking-skill development?*

We identify 4 dyadic peer-influence factors (context-driven feedback, first-hand demonstration, empathetic support and acknowledgment) and 3 community-level factors (diversity of perspectives, sense of community and comfort in interaction). These factors can potentially open up further avenues for affective explorations in the speaking-skills training context. Thus, our contributions are summarized as follows:

- Showing that improvement suggestions and peers' performances add incremental value to one's performance;
- Classifying tips from non-tips using linguistic features;
- Identifying key dyadic and network-level peer-interaction factors that contribute to speaking-skills development.

II. Background

A. Affective Computing in Speaking-skills Development

A large body of work has focused on sensing and evaluating various aspects of speaking-skills. Using multimodal cues, researchers have attempted to evaluate public speaking and job interview performances [17]–[21]. Progress in automatic sensing of verbal and non-verbal behaviors have led to the development of computational frameworks to objectively quantify feedback. For example, Logue helps users become aware of their non-verbal behaviors [22]. Rhema uses Google Glasses to give feedback on speech rate and volume [23]. AutoManner extracts idiosyncratic mannerisms from speech videos to draw the users' attention to those [6]. AwareMe uses a wearable device to generate feedback on pitch, use of filler words and speech rate [24]. MACH's 3D virtual coach allows anyone to practice job interviews, providing post-session feedback on verbal and non-verbal speaking attributes [9]. Virtual audience simulation has been used as part of live feedback systems [25] and in assessing speaking anxiety [26]. Most of these prior systems were designed to give automated feedback to an individual. Yet, using affective tools to sense and intelligently assist speaking-skill related feedback exchanges in a larger-scale community setting has received little research attention. We contribute to this scope.

B. Peer-influence in Learning

Peer-influence on one's learning process is well researched in education literature [15]. The knowledge build-up process is known to be facilitated by peer discussion and information sharing in small groups [27]. Such interactions help both the 'tutor' and 'tutee'. The tutee benefits from the knowledge gathered from the interactions [28]. The tutor also benefits from the meta-cognitive task of explaining something, which helps to clarify his/her own understanding [29], [30]. Some debate does exist on the effectiveness of feedback interventions, as mixed outcomes have been reported in literature [31]. Development of a sense of community [32], comparison of performances [33] and encouragement [33] are known to impact learning outcomes, among other factors. Mackness *et al.* found autonomy, diversity, openness, connectedness and interactivity to impact the learning experiences of MOOC participants [34]. In the more relevant domain of soft-skills, network effects have recently been shown to help develop teamwork [35] and speaking-skills [12]. These findings encourage our exploration of speaking-skills development from a peer-learning standpoint. Furthermore, one can argue that the highly tacit components of developing speaking-skills can be aided by observation, imitation and practice [36]. This in turn can benefit from watching the peer-learners perform and progress themselves. Such use of the video modality in peer-induced skill development has received little attention in literature. In this paper, we address this gap.

III. Dataset

A. Online Interaction Data

We hired 60 participants (28 male, 32 female) from Amazon Mechanical Turk. They were aged between 18-54 (18y-24y: 2, 25y-34y: 29, 35y-44y: 18, 45y-54y: 11) and 86.7% of them had an education level beyond high school (bachelor's: 24, master's: 11, Ph.D.: 1). We gave them 5 common job interview prompts in 10 days, one every other day. The prompts were: *(1) Tell me about yourself, (2) Describe your biggest weakness, (3) Tell me about your greatest achievement, (4) Describe a conflict or challenge you faced, and (5) Tell me about yourself.* The first prompt was repeated as the fifth for comparison purposes. The participants video-recorded their responses to the prompts using webcams ($\sim$ 2 mins in duration, with options for re-recording). They also gave comments and performance ratings to the videos of 3 peers they were connected with. The peers returned the favor. These peer-connections were made by randomly assigning the participants to the nodes of the network structure shown in Figure 1b. Two independent instances of this 30-person structure were used to make the data collection feasible. The node assignments remained fixed throughout the study, and the participants could interact only with the assigned peers. The participants were instructed to focus their comments and ratings on 3 attributes: appropriate uses of (i) smiles, (ii) volume modulation and (iii) gestures. To each peer's video, the participants were required to give at least 3 comments (on the three skill-attributes) and 4 ratings (on the three skill-attributes and an overall rating; on 5-point scales). They tagged their comments with the respective skill-attributes, using an 'other' tag for any miscellaneous comment. The comments were privately viewable only by the receivers, to narrow down their sources of peer-learning. The participants filled up pre- and post-study surveys, and were paid \$25 upon completion of all the tasks. The study was conducted with approval from IRB. The dataset contains 305 videos, 2857 comments and 3660 ratings in total.

B. Human Annotation

Most of the comments in the dataset focused on praising and encouraging (*"Good job!"*), while some provided directions for further improvement (*"You may want to use gestures slightly more."*). Two RAs annotated each comment on whether it contains suggestions for improvement or not ($1/0$ labels). The initial inter-rater agreement (Cohen's κ) on annotating 2857 comments was 0.88. The disagreements were then resolved through discussion. Finally, 627 of the comments were labeled to contain improvement suggestions (label 1). For brevity, we refer to these comments as 'tips' in the sequel.

IV. Experiments

The participants had two major sources of peer-influence: receiving feedback comments and observing the peers' performances. To answer our first research question (**R1**), we analyze how these sources correspond to performance improvement.[1]

A. Effects of Comment-based Feedback

The 'tips' annotation allows us to study the role of improvement suggestions. We denote the set of skill attributes as $S = \{$(appropriate uses of) Smile, Volume modulation, Gesture$\}$. The performance metric, $r_{i,s}^{(p)}$, is defined as the average rating received by participant i on skill attribute $s \in S$ in prompt $p \in \{1, 2, ..., 5\}$. Similarly, the number of tips received by i in the respective s and p is denoted with $n_{i,s}^{(p)}$.

We compare ratings among videos that received $n_{i,s}^{(p)} = 0, 1$ and ≥ 2 tips, for $s \in S$ and $p \in \{1, 2, 3, 4\}$. The comparisons are shown in Figure 2, along with the t-test and Cohen's d results for brevity. The purple and green boxplots respectively correspond to ratings in the same (p) and following $(p + 1)$ prompt as the received comment. The purple boxplots show that videos receiving no tip on a certain skill attribute were rated significantly higher than videos with a single tip [Fig. 2(i)]. Again, videos receiving 1 tip were rated significantly higher than those with ≥ 2 tips [Fig. 2(ii)]. This is intuitive, as one would expect the lower performing videos to receive more improvement tips from the peers than the better rated ones. In fact, $n_{i,s}^{(p)}$ is negatively correlated with $r_{i,s}^{(p)}$, with Pearson corr. coeff.$= -0.38$, $p < 0.001$.

In the prompt *after* receiving the comment (green boxplots), the ratings improved significantly for users who had received 1 and ≥ 2 tips [Fig. 2(vi) and Fig. 2(vii)], but no significant change happened for the 0 tip case [Fig. 2(v)]. In other words,

[1]Data and analysis code: https://github.com/ROC-HCI/Upskilling_together

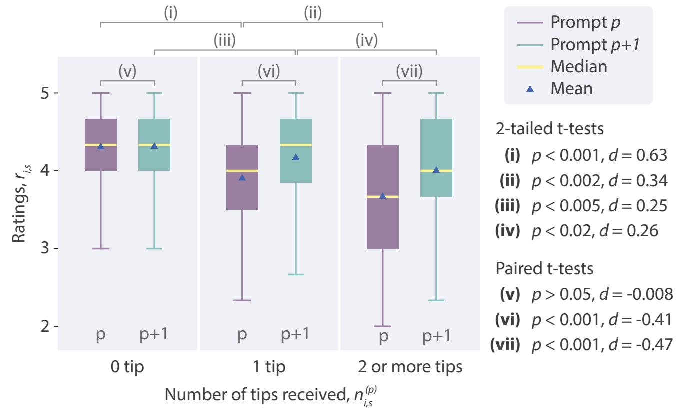

Fig. 2. Ratings vs number of tips received. Purple and green plots respectively correspond to ratings in the same and following prompt as the comment.

the tip receivers caught up with those who received no tip. There still remained significant rating differences between 0 tip and 1 tip cases [Fig. 2(iii)]; and between 1 tip and ≥ 2 tip cases [Fig. 2(iv)]—although with decreased effect sizes. These results show a systematic association between receiving tips and improving performances in the given skill attributes.

B. Effects of Peers' Performances

Each participant in the dataset was connected with 3 fixed peers throughout the study. This static network condition allows us to study the correspondence between a participant's own performance and his/her peers' performances across prompts. We take the average of one's peers' ratings in a given skill attribute as the 'peer exposure' metric. Namely, the peer exposure for participant i is taken as $\bar{r'}_{i,s}^{(p)} = mean(r_{j,s}^{(p)})$, for all j in i's neighborhood, $s \in S$ and $p \in \{1, 2, 3, 4\}$.

Statistical tests show that when the average of peers' ratings is worse than the participant's own rating, the participant's performance suffers in the following prompt and has a small yet significant decrease in ratings [Fig. 3(i)]. On the other hand, when the average of one's peers' ratings (i.e., peer exposure metric) is better than one's own ratings, his/her improvement in the following prompt is significant with a larger effect size [Fig. 3(ii)]. These results indicate a systematic correspondence between the user's and his/her peers' aggregated performances.

C. Affective Characteristics of Tips

We use linguistic analysis to infer basic emotions, social personality and sentiment ('Affective features') from the written comments. In particular, we use IBM Watson Tone Analyzer [37], [38] to extract comment-wise scores on Emotion (*anger, disgust, fear, joy* and *sadness*), Language (*analytical, confident* and *tentative*), Social Personality (*openness, extraversion, emotional range, conscientiousness* and *agreeableness*) and Sentiment (*sentiment score*). We also extract sentiment scores from Vader [39] (*neutral, positive* and *compound*) for each comment.

We run hypothesis tests to explore the differences in affective features between tips and non-tips. The results are

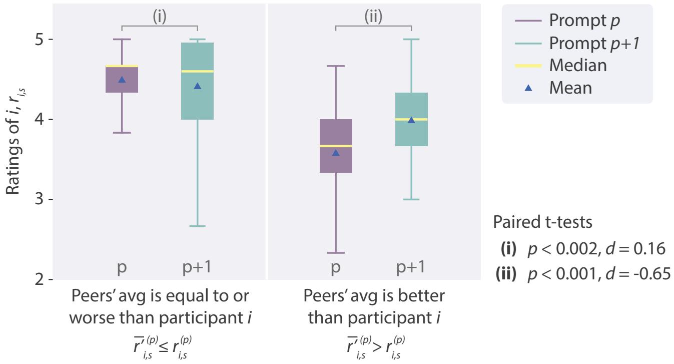

Fig. 3. Ratings when the average of one's peers' ratings is equal or worse than his/her own rating vs when the average is better than one's own rating.

TABLE I
T-TEST RESULTS BETWEEN TONES OF TIPS AND NON-TIPS (d=COHEN'S D)

Tone	p-val	d	Tone	p-val	d
neutral	<.001	0.87	conscientious	<.001	-0.41
tentative	<.001	0.79	compound	<.001	-0.46
fear	<.001	0.70	confident	<.001	-0.51
sadness	<.001	0.61	joy	<.001	-0.67
disgust	<.001	0.39	extraversion	<.001	-0.83
anger	<.001	0.26	positive	<.001	-0.91
emotional range	<.001	-0.28	sentiment score	<.001	-0.95

shown in Table I. The tips have significantly higher scores in *neutral*, *tentative*, *fear*, *sadness*, *disgust* and *anger* than non-tips. These tone categories have a non-positive connotation—which is intuitive, as the tips point out issues that can be improved. On the other hand, the tips have significantly lower scores in *sentiment score*, *positive*, *extraversion*, *joy*, *confident*, *compound*, *conscientiousness* and *emotional range*—which are mostly positive tones. To account for the number of features used, all of the reported p-values are Bonferroni corrected (p-val×17). The features *analytical*, *openness* and *agreeableness* do not show statistically significant results.

To understand how well the affective features characterize the tips, we formulate a classification problem where the binary labels (tip/not a tip) of the comments are predicted by the 17 affective features. Since there are 627 tips among 2857 total comments, we randomly sample 627 non-tip comments to create a balanced dataset. We use 4 algorithms: Logistic Regression, Linear Support Vector Machine (SVM), Kernel SVM (radial basis function kernel) and Random Forest. The dataset is split $7:3$ into training and test sets. We tune the hyper-parameters (e.g., the slack parameter C in SVM) by randomized search with 3-fold cross validation. The test-set results are summarized in Table II. As can be seen, all of the algorithms perform well, with Logistic Regression giving the best AUC (0.89) and Random Forest giving the best F1-score (0.81) and accuracy (0.82). Feature selection using Logistic Regression (with L1 regularization), Random Forest and Linear SVM consistently shows 5 of the features to be most predictive of the tips' characteristics: *tentative*, *extraversion*, *sentiment score*, *positive* and *neutral*. This is

TABLE II
TEST-SET RESULTS OF TIPS VS NON-TIPS CLASSIFICATION

	AUC	F1-score	Accuracy
Linear SVM	0.88	0.80	0.81
Logistic Regression	**0.89**	0.80	0.80
Random Forest	0.87	**0.81**	**0.82**
RBF SVM	0.88	0.79	0.80

intuitive and powerful: while the comments with suggestions were naturally less positive, less extraverted and had more of a neutral tone, the peers were at the same time being tentative and measured in communicating their suggestions.

The above analysis shows that the affective features adequately characterize the tips and enable classifying them from non-tips, as we sought to understand in the second research question (**R2**). This opens up the possibility of using the affective features as a sensing mechanism for the suggestive content in the comments, sidestepping the need for tips annotation.

D. Predicting Future Ratings

In the previous sections we exercised caution not to solely attribute the rating improvements to the two sources of peer-influence that we studied separately. This is because the statistical tests can be prone to confounding effects, making it difficult to decouple the individual effects of the two sources. In this section, we shed light on the incremental values added by the two sources of peer learning by approaching the problem as a prediction formulation.

For a skill attribute $s \in S$, we specify a linear model for the performance rating of individual i in prompt $p+1$ as,

$$r^{(p+1)}_{i,s} = \beta_0 + \beta_1 r^{(p)}_{i,s} + \beta_2 \bar{r}'^{(p)}_{i,s} + \sum_k \eta_k \bar{a}^{(p)}_{i,s,k} \qquad (1)$$

Here, $r^{(p)}_{i,s}$ is the individual's own rating in prompt $p \in \{1,2,3,4\}$, and acts as a baseline predictor of the future rating. $\bar{r}'^{(p)}_{i,s}$ is the peer exposure metric, which captures the effects of observing or learning from the peers' video performances. $\bar{a}^{(p)}_{i,s,k}$ is the average tone of the comments received by i in the tone category k, where $k \in$\{17 affective features\} as used previously. These tone features capture the suggestive content of the incoming comments, as described earlier.

We use Ridge regression with L2 regularization, LASSO regression with L1 regularization, and Support Vector Regression (SVR) to predict the future ratings using Eq. 1. Once again, we use a $7:3$ training and test split, and use randomized search with 3-fold cross validation to tune the hyper-parameters. The test-set results are summarized in Table III. As can be seen from the table, as the tone features and the peers' average rating features are added incrementally, there is an improvement in the model's prediction performance. The correlation coefficients of the predicted ratings and the ground-truth ratings increase, with all the reported correlation coefficients having $p < 0.001$. This p value is computed against a null hypothesis of no correlation, indicating that the correlation coefficients presented here are systematic. This

TABLE III
TEST-SET RESULTS OF FUTURE RATINGS PREDICTION

Input features	Ridge	SVR	LASSO
	Corr. (p-val)	Corr. (p-val)	Corr. (p-val)
Current rating	0.585 (<.001)	0.585 (<.001)	0.585 (<.001)
+ tone features	0.593 (<.001)	0.597 (<.001)	0.585 (<.001)
+ peers' avg. ratings	0.604 (<.001)	0.608 (<.001)	0.596 (<.001)

provides validation that both the received comments and peers' performances concurrently influence one's skill development.

V. QUALITATIVE ANALYSIS

A. Method

We use grounded theory [40] to analyze the exchanged comments and the post-study survey responses. We use selective coding and analysis of this data to understand the peer-influence factors, towards answering the third research question (**R3**). First, we flag all instances where dyadic or community-level influence factors come up. We then cluster them into conceptual categories. Iterative coding is used to refine the outcomes, which results in 7 conceptual categories. Further reflection suggests that 4 of those operate at an individual (dyadic) level, and 3 at a community-level.

B. Results

In the pre- and post-study surveys, the participants reported their confidence in speaking-skills on a 1-7 Likert scale. The improvement was significant ($p < 0.01$ in Mann-Whitney U test). The first and fifth prompts were the same, allowing us to compare the 'overall' ratings received by the users, which also improved significantly ($p < 0.01$ in paired t-test). We identified the following peer-influence factors:

(1) Dyadic peer-learning factors

(a) Exchanging context-driven feedback. In the post-study survey, the participants repeatedly mentioned benefiting from the context-driven and personalized comments. They reported that the feedback revealed useful insights on their speaking behavior (Mean=5.85, SD=1.23 on a 1-7 Likert scale). They further mentioned both giving and receiving feedback comments to be helpful.

Case Study. In prompt 1, ID6 received the following feedback, *"Give your personality a chance to shine through. Smile and allow the audience to see that you are excited."* In the following prompt, the attempt to address the feedback was appreciated by ID6's peers, *"You are much more relaxed in this video. You are very relatable and seem to do well with relating to your audience."* Finally, in the fifth prompt, the improvements won enthusiastic comments from the peers, *"I love that your personality is really shining through. Great job!"* This case study corroborates our quantitative insights on the effectiveness of tips.

(b) Demonstrating how it is done. The participants reported in the survey how watching videos of peers helped them pick ideas: *"It was very helpful to see how others did, and to take tips from what you liked from other people's videos."*

Case Study. ID18 used minimal gestures in the first prompt—it is naturally harder to use hand and body gestures in a webcam interview setting. However, his peer, ID9, positioned the camera slightly farther away from herself so that she could stand up and use gestures more animatedly. She shared the idea with ID18: *"It's a bit unusual to stand in front of your computer, but if you can find a way to do so comfortably, I believe it would help you with your movement and gestures."* ID18 could see how ID9 herself implemented the idea in her video. By the fourth prompt, ID18's use of gestures appeared natural and fluent, as appreciated by a peer: *"You're doing great with appropriate gestures. Really goes along with your speech and helps bring more enthusiasm and connectivity with your audience."* (https://youtu.be/AYcH0g4v8to) This corroborates our findings on the utility of the video modality.

(c) Extending empathetic support. The experiment demography was diverse in its age-range and professional backgrounds. This allowed us to observe cases where empathy from peers played a role in bringing out the best in a participant.

Case Study. ID7 was a stay-at-home mom, outside the job market at the time of the study. In the third prompt, "Tell me about your greatest achievement", she talked about raising her children, instead of any professional achievement like others. She had a defensive tone in doing so, which was quickly addressed by her peers: *"The sentence drops down in tone at the end of the sentences where you describe yourself as a stay-at-home mom. Almost like you're not impressed with yourself... You have pride in what you do and it comes through later on in the video, but start it off good too! This is what you do! You do it well! Project it in your tone to convince us!"* ID7's persona was dramatically different in the following prompt, as a peer points out: *"You came out so confident in the beginning and I felt that throughout your video. This is probably your strongest video yet from what I've seen."* (https://youtu.be/AYcH0g4v8to) This shows how non-technical yet empathetic support can play a role in shaping people's confidence in speaking.

(d) Acknowledging progress and encouraging one another. A large portion of the exchanged comments involved acknowledging improvement and encouraging the peers. While encouragement is known to help in keeping the learners engaged [33], [41], some users found the lack of improvement pointers uninviting, *"I tried to give constructive feedback, I didn't feel like I received much. It was mostly praise and people being nice when I could have really used some advice."* Most of these comments got annotated as non-tips (label 0), and participants receiving no tip showed no significant improvement in their following prompts.

(2) Peer-learning factors at a community level

(a) Bringing in diverse perspectives. Having multiple peers allowed the participants to benefit from the diversity of opinions and insights: *"It was interesting to get opinions* *from such a varied group of people. They point out things you might not otherwise think of."* Such aggregation of peer opinions can allow a user to get an insightful picture, despite the peers being non-experts or learner themselves.

(b) Developing a sense of community. Despite the number of peers being low, the participants reported feeling as a part of a community, which helped build trust and comfort in the interactions: *"Receiving genuine feedback from other people who were also trying to improve was great. I felt like we were in it together, it felt less intimidating."* This collaboration factor was also reflected in the survey responses to whether they felt that they were part of a community (Mean=6.38, SD=0.87). The act of speaking often intimidates people, and the sense of community helped alleviate that to some extent.

(c) Creating a comfortable and enjoyable interaction experience. The participants reported enjoying watching others' videos and giving feedback to them (Mean=6.33, SD=0.83). They also reported that they felt comfortable uploading their own videos for feedback (Mean=6.32, SD=1.08). One user wrote, *"I thought it was kind of fun to go through the video creation process. Getting to know some of the other participants was also enjoyable."* In particular, this enjoyment was reported in a community context, where having multiple peers contributed to creating a merry experience.

VI. Discussion

We collected a temporal dataset comprising videos, peer comments and ratings. Since our objective was to study the peer-learning components in speaking-skills development, we needed to minimize noises coming from non-peer sources of learning. Using a static network made the set of peers fixed for everyone, so the sources of peer-learning could be narrowed down. Making the incoming comments privately viewable limited the participants' exposure to exchanges in other people's videos, further limiting the sources of learning. While these laboratory settings helped us study the peer effects in a cleaner manner, a real-world interaction platform may benefit from dynamic interaction and open discussions.

We characterized comments that offer improvement suggestions using linguistic features, and showed that the features can classify whether or not a comment includes improvement suggestions (AUC 0.89). Such features have previously been used in literature [17]. This finding can help peer-learning systems to have a high-level understanding of the exchanges, creating intelligent intervention possibilities.

Watching the peers' videos to give feedback can influence one's skill development two key ways: by picking up ideas from the first-hand demonstrations by the peers, and by comparison of performances [33]. We identified survey responses and case studies where participants incorporated skill attributes from their peers into their own future videos. Quantitative analysis showed that having better peers corresponded to performance improvement, while having worse peers actually led to a slight drop in performance. While video-based interaction is common in social media (e.g., Snapchat), its effects are less explored in peer-learning and skill development literature. We shed light on its value.

We used a linear prediction model to show that a participant's future rating prediction is incrementally benefited as features from comment-tones and peer-ratings are added. The point was to illustrate that the signals from the two modalities of interaction have predictive information in them, as seen from the incrementally better predictions. Naturally, these features do not cover all the possible factors that influence performance, which is why the prediction results are not perfect. Similar linear prediction models have previously been used in behavior contagion analysis in temporal network data [42]—although our model is simpler to avoid overfitting.

Using a ground up qualitative approach, we identified dyadic and network-level peer-learning factors. Some of these factors were reported in education literature on networked learning. For example, receiving constructive feedback [33], [41], diversity of opinion [34], sense of community [43], acknowledgment and encouragement [33], [41] were previously reported, as we corroborated for speaking-skills development. These provide actionable insights for the development of training systems on speaking-skills, and motivates the embracing of such human elements in the training loop.

Our study is not without limitations. A small-scale experiment with paid participants led to the absence of anti-social behavior in our data. However, real-life online platforms often suffer from anti-social bullying [44] and come with privacy concerns. These need careful considerations in the interaction design. The generalizability of our results can be confirmed by exploring other oral communication scenarios such as doctor-patient communication. These are part of our future work.

VII. Conclusion

We explored speaking-skills development from a peer-learning standpoint. Using a temporal dataset with bidirectional interaction benefits, we showed that receiving improvement suggestions and having access to better peers both have systematic associations with performance improvement. We found that linguistic features enable the classification of tips from non-tip comments. Qualitative analysis revealed 4 dyadic and 3 community-level peer-influence factors. These insights inform both the human and AI-powered ends of the speaking-skill training spectrum. Research on automated tools can benefit from a deeper understanding of how human peers generate effective feedback. The identified factors and associated characterizations can also allow the development of intelligent interaction systems that compliment and enhance human capabilities of subjective judgment in an informed way.

Perhaps, it may be possible in the near future for computers to generate feedback that embraces the human elements of social interaction. How machine feedback may compare to feedback coming from a human would pose a very interesting research question. We believe our findings in this paper may instigate an important dialogue of whether we indeed want to pursue research that could potentially replace the power of what human collaboration can enable.

REFERENCES

[1] T. T. H. Truong, R. S. Laura, and K. Shaw, "The importance of developing soft skill sets for the employability of business graduates in Vietnam: A field study on selected business employers," *Journal of Education and Culture Studies*, vol. 2, no. 1, p. 32, 2018.

[2] P. C. Kyllonen, "Soft skills for the workplace," *Change: The Magazine of Higher Learning*, vol. 45, no. 6, pp. 16–23, 2013.

[3] M. M. Robles, "Executive perceptions of the top 10 soft skills needed in today's workplace," *Business Communication Quarterly*, vol. 75, no. 4, pp. 453–465, 2012.

[4] Toastmasters. (2019, Retrieved 17 April) Toastmasters - where leaders are made. https://www.toastmasters.org/. Toastmasters.

[5] O. T. Room. (2019, Retrieved 17 April) Own the room - unleashing human potential. http://www.owntheroom.com/. Own The Room.

[6] M. I. Tanveer, R. Zhao, K. Chen, Z. Tiet, and M. E. Hoque, "Automanner: An automated interface for making public speakers aware of their mannerisms," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 2016, pp. 385–396.

[7] M. Fung, Y. Jin, R. Zhao, and M. E. Hoque, "ROC Speak: Semi-automated personalized feedback on nonverbal behavior from recorded videos," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 1167–1178.

[8] R. Zhao, V. Li, H. Barbosa, G. Ghoshal, and M. E. Hoque, "Semi-automated & collaborative online training module for improving communication skills," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, p. 32, 2017.

[9] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "MACH: My automated conversation coach," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2013, pp. 697–706.

[10] M. E. Hoque and R. Picard, "Rich nonverbal sensing technology for automated social skills training," *Computer*, vol. 47, pp. 28–35, 2014.

[11] F.-F. Li. (2018) How to make A.I. that's good for people. https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html.

[12] R. Shafipour, R. A. Baten, M. K. Hasan, G. Ghoshal, G. Mateos, and M. E. Hoque, "Buildup of speaking-skills in an online learning community: A network-analytic exploration," *Palgrave Communications*, vol. 4, no. 1, pp. 1–10, 2018.

[13] C. Jones, *Networked Learning: An Educational Paradigm for the Age of Digital Networks*. Springer, 2015.

[14] M. Dado and D. Bodemer, "A review of methodological applications of social network analysis in computer-supported collaborative learning," *Educational Research Review*, vol. 22, pp. 159–180, 2017.

[15] K. Topping, C. Buchs, D. Duran, and H. Van Keer, *Effective Peer Learning: From Principles to Practical Implementation*. Taylor & Francis, 2017.

[16] L. Harasim, *Learning Theory and Online Technologies*. Taylor & Francis, 2017.

[17] M. I. Tanveer, S. Samrose, R. A. Baten, and M. E. Hoque, "Awe the audience: How the narrative trajectories affect audience perception in public speaking," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, p. 24.

[18] L. Chen, R. Zhao, C. W. Leong, B. Lehman, G. Feng, and M. E. Hoque, "Automated video interview judgment on a large-sized corpus collected online," in *7th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 504–509.

[19] L. Chen, C. W. Leong, G. Feng, C. M. Lee, and S. Somasundaran, "Utilizing multimodal cues to automatically evaluate public speaking performance," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2015, pp. 394–400.

[20] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated analysis and prediction of job interview performance," *IEEE Transactions on Affective Computing*, 2016.

[21] M. R. Ali, F. Ciancio, R. Zhao, I. Naim, and M. E. Hoque, "ROC Comment: Automated descriptive and subjective captioning of behavioral videos," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 928–933.

[22] I. Damian, C. S. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André, "Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques," in *Proceedings of the 33rd annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 565–574.

[23] M. I. Tanveer, E. Lin, and M. E. Hoque, "Rhema: A real-time in-situ intelligent interface to help people with public speaking," in *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 2015, pp. 286–295.

[24] M. Bubel, R. Jiang, C. H. Lee, W. Shi, and A. Tse, "AwareMe: Addressing fear of public speech through awareness," in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2016, pp. 68–73.

[25] L. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer, "Cicero - Towards a multimodal virtual audience platform for public speaking training," in *International Workshop on Intelligent Virtual Agents*. Springer, 2013, pp. 116–128.

[26] T. Wörtwein, L.-P. Morency, and S. Scherer, "Automatic assessment and analysis of public speaking anxiety: A virtual audience case study," in *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2015, pp. 187–193.

[27] N. Davidson and C. H. Major, "Boundary crossings: Cooperative learning, collaborative learning, and problem-based learning." *Journal on Excellence in College Teaching*, vol. 25, 2014.

[28] R. Ferguson, "Peer interaction: The experience of distance students at university level," *Journal of Computer Assisted Learning*, vol. 26, no. 6, pp. 574–584, 2010.

[29] L. De Backer, H. Van Keer, and M. Valcke, "Exploring evolutions in reciprocal peer tutoring groups' socially shared metacognitive regulation and identifying its metacognitive correlates," *Learning and Instruction*, vol. 38, pp. 63–78, 2015.

[30] R. D. Roscoe and M. T. Chi, "Tutor learning: The role of explaining and responding to questions," *Instructional Science*, vol. 36, no. 4, pp. 321–350, 2008.

[31] A. N. Kluger and A. DeNisi, "The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory," *Psychological Bulletin*, vol. 119, no. 2, p. 254, 1996.

[32] X. Liu, R. J. Magjuka, C. J. Bonk, and S.-h. Lee, "Does sense of community matter? An examination of participants' perceptions of building learning communities in online courses," *Quarterly Review of Distance Education*, vol. 8, no. 1, p. 9, 2007.

[33] H. Oinas-Kukkonen and M. Harjumaa, "Persuasive systems design: Key issues, process model and system features," in *Routledge Handbook of Policy Design*. Routledge, 2018, pp. 105–123.

[34] J. Mackness, S. Mak, and R. Williams, "The ideals and reality of participating in a MOOC," in *Proceedings of the 7th International Conference on Networked Learning 2010*. University of Lancaster, 2010.

[35] Á. Fidalgo-Blanco, M. L. Sein-Echaluce, F. J. García-Peñalvo, and M. Á. Conde, "Using learning analytics to improve teamwork assessment," *Computers in Human Behavior*, vol. 47, pp. 149–156, 2015.

[36] P. Busch, *Tacit Knowledge in Organizational Learning*. IGI Global, 2008.

[37] IBM. (2019, April) Tone analyzer: Understand emotions and communication style in text. https://www.ibm.com/watson/services/tone-analyzer/. IBM.

[38] ——. (2019, April) The science behind the service. https://cloud.ibm.com/docs/services/tone-analyzer?topic=tone-analyzer-ssbts#ssbts/. IBM.

[39] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[40] J. Corbin and A. Strauss, *Basics of Qualitative Research*. Sage Publications, Inc, 2014.

[41] X. Wang, D. Yang, M. Wen, K. Koedinger, and C. P. Rosé, "Investigating how student's cognitive behavior in mooc discussion forums affect learning gains." *International Educational Data Mining Society*, 2015.

[42] S. Aral and C. Nicolaides, "Exercise contagion in a global social network," *Nature Communications*, vol. 8, p. 14753, 2017.

[43] N. Luo, M. Zhang, and D. Qi, "Effects of different interactions on students' sense of community in e-learning environment," *Computers & Education*, vol. 115, pp. 153–160, 2017.

[44] J. Cheng, M. Bernstein, C. Danescu-Niculescu-Mizil, and J. Leskovec, "Anyone can become a troll: Causes of trolling behavior in online discussions," in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2017, pp. 1217–1230.